

Sociology 2275, Social Network Analysis, Fall, 2022

Assignment 6: Structurally Cohesive Subgroups

This assignment provides a little experience in working with some tools for identifying “structurally cohesive subgroups” and “community detection.” Software notes on “xUCINET Functions for Identifying Clusters, Cliques, Communities and the like” and “Hierarchical Clustering in xUCINET” will be helpful in completing it, we hope.

For this exercise, again work with the data on an undirected (as in assignment 5) network of discussions among executives in “The Corporation.”

1. Enumerate the membership of cliques in the undirected network. **Report both the total number of cliques (of size 3 or greater), and the executives who are in the largest one** (if there is more than one largest clique, report the membership of all cliques of that size).
2. Group the executives into larger clusters on the basis of hierarchical clustering of their shared memberships in cliques. **Report:**

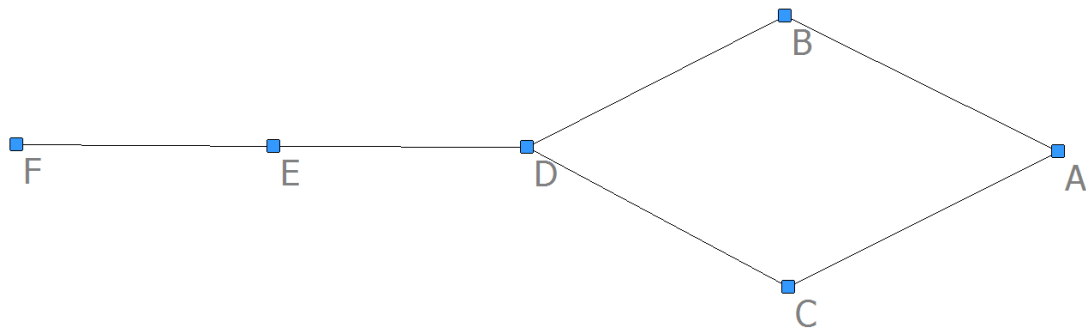
- a. the number of clusters you select
- b. the membership of each cluster
- c. a subgroup density table illustrating the outcome of this analysis

3. The Girvan-Newman algorithm for community detection relies on a centrality measure for *edges* known as edge betweenness. An edge’s shortest-path betweenness is defined as the number of times it lies on geodesics connecting pairs of vertices:

$$C_B(x_{km}) = \sum_{i=1}^N \sum_{j=i+1}^N b_{ij}(x_{km}) = \sum_{i=1}^N \sum_{j=i+1}^N \frac{g_{ij}(x_{km})}{g_{ij}},$$

where $g_{ij}(x_{km})$ is the number of times an edge linking vertices k and m lies on a geodesic linking vertices i and j , and g_{ij} is the total number of geodesics linking those vertices.

Using the small network studied in assignment 4,



calculate edge betweenness scores for edges B-D, C-D, and E-D (the betweenness of edges A-B, A-C, and E-F is lower than for these three). **Which edge has the greatest betweenness**, that is, which would be the first to be deleted by the Girvan-Newman algorithm?

4. Identify “communities” in the undirected Corporation network using the Girvan-Newman algorithm. **Report the membership of the communities you identify, the modularity value for the partition, and a density table illustrating the relationships within and between communities.**

5. Using the attribute data about the executives, try to assign an interpretation to the subgroups you identified in part 4. Use simple cross-tabulations or tables of means. **What do the executives mapped into particular communities share, in terms of seniority, office, and project experience?**

Due: Thursday, October 27.

Note: These assignments will not be graded, but we will take note of whether or not you do them. Please submit to either derick_baum@g.harvard.edu or pvm@wjh.harvard.edu .